

## Posudek vedoucího diplomové práce

Jméno posuzovatele: David Hoksza

Autor práce: Kateřina Břicháčková

Název práce: Využití anotací primární struktury pro strukturní predikci protein-ligand aktivních míst

Cílem práce bylo vytvořit softwarový framework umožňující analýzu anotací residuí proteinu s cílem odhalit jejich potenciál pro strukturní predikci protein-ligand aktivních míst se zaměřením na využití nástrojem P2Rank. Téma vyžadovalo seznámení se s oblastí detekcí protein-ligand anotací, softwarovými nástroji, které jsou v této oblasti používány, zvláště pak s nástrojem P2Rank. Dále pak bylo třeba navrhnout experimenty, které by umožnili testovat potenciální důležitost vlastností pro detekci interakčních míst. Následně, bylo třeba vytvořit softwarový framework implementující navrženou metodologii a tuto vyhodnotit jak na úrovni důležitosti vlastností obecně, tak ověření jejich vhodnosti pro využití konkrétně v metodě P2rank. Samotná textová část práce pak reflektuje tyto kroky.

Rešeršní část práce obsahuje krátký přehled stávajících směrů v predikci protein-ligand interakcí. Více prostoru je pak dáno popisu fungování nástroje P2Rank a způsobu vyhodnocení kvality predikcí v této oblasti, který se liší od standardních metrik používaných v podobných predikčních úlohách.

Sekce popisující metodologii testování obsahuje popis mapování sekvenčních a strukturních dat a způsob získání anotací různých typů z různých zdrojů, které jsou dále použity při evaluaci. Velká pozornost je pak věnována popisu statistických testů použitých pro vyhodnocení potenciální důležitosti konkrétní anotace v závislosti na jejím typu (spojitá vs. diskrétní veličina). Popis jednotlivých statistických testů je proložen popisem jejich implementace ve výsledném softwarovém frameworku. Tato sekce nejenže vysvětluje, jak probíhá testování, ale taky popisuje různá úskalí spojená s používáním p-value při velkém množství dat a nutnost použití velikosti efektu k vzájemnému porovnání vlivu jednotlivých anotací. Tyto aspekty jsou pak testovány v sekci věnované evaluaci frameworku.

Samotná evaluace pak probíhá na datových sadách odvozených z těch, které jsou využity pro hodnocení predikčních schopností nástroje P2Rank. Výstupem je jak hodnocení potenciálního vlivu jednotlivých typů anotací na predikci, tak následné vyhodnocení aktuálního dopadu. Pozitivně hodnotím, že výsledky nejsou čtenáři pouze předkládány, ale jsou důkladně diskutovány jak z pohledu statistického, tak z pohledu biologického. Na závěr je metodologie podrobena aplikaci na případu vyhodnocení potenciálního vlivu nové metody výpočtu konzervovanosti. Zde je třeba podotknout, že tato aplikace se objevila až po dokončení vývoje frameworku, a tudíž ji lze brát jako nezávislé vyhodnocení aplikovatelnosti navrženého řešení.

Kvalitu textové práce, která byla vypracována v anglickém jazyce, hodnotím pozitivně, a to jak z pohledu obsahu, tak z pohledu čtivosti.

Nezávisle je třeba u tohoto typu práce hodnotit i softwarovou část, kterou je třeba, podle mého názoru, v oboru bioinformatiky klást prakticky na roveň textové části. V tomto ohledu hodnotím práci t06 pozitivně. Softwarový kód je dobře dokumentovaný, ať už z pohledu vývojového nebo z pohledu

uživatelské dokumentace. Celý projekt je pak zveřejněn na platformě GitHub, což umožňuje její další bezproblémový vývoj.

Kladně pak také hodnotím přístup studentky k práci na projektu. Pravidelně jsme byli v kontaktu a řešili dílčí úlohy a problémy. Typicky, když studentka narazila na problém, tak nejdříve nastudovala literaturu, případně další zdroje, řešící daný problém a se mnou, jako s vedoucím, pak spíše konzultovala už nastudované alternativy řešení daného problému. To se týkalo jak softwarových problémů tak problémů týkajících se např. statistického testování. Aktivně řešila i problémy s nástroji a daty třetích stran jako byly nalezené problémy v BioPythonu nebo problémy s novými rozhraními PDBe-KB.

S ohledem na výše uvedené skutečnosti nemohu jinak, než doporučit diplomovou práci k obhajobě.